

Football Match Result Prediction - Project Protocol

Executive Summary

This project uses machine learning to predict football match outcomes (Home Win, Draw, Away Win) from match statistics. Using an XGBoost classifier, we achieved 99% accuracy on the test set.

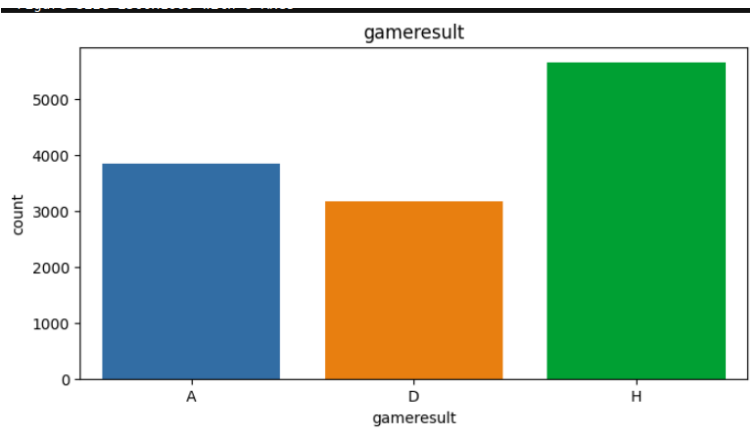
1. Introduction

Dataset from: <https://www.kaggle.com/datasets/technika148/football-database>. Goal is to predict match result: Home Win, Draw, or Away Win.

2. Objective

Create a model to classify football match outcomes based on match stats.

3. Dataset Overview



4. Data Journey

Football Match Result Prediction - Project Protocol

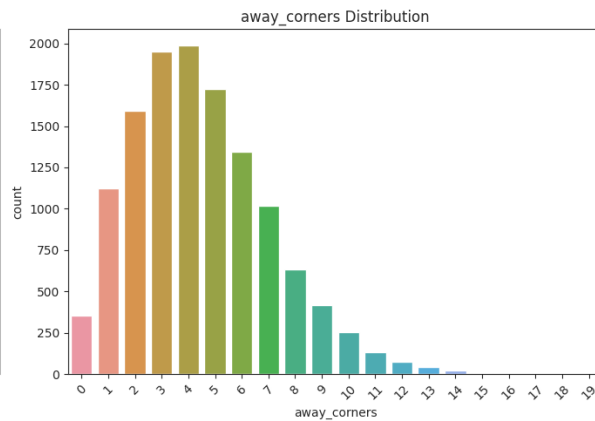
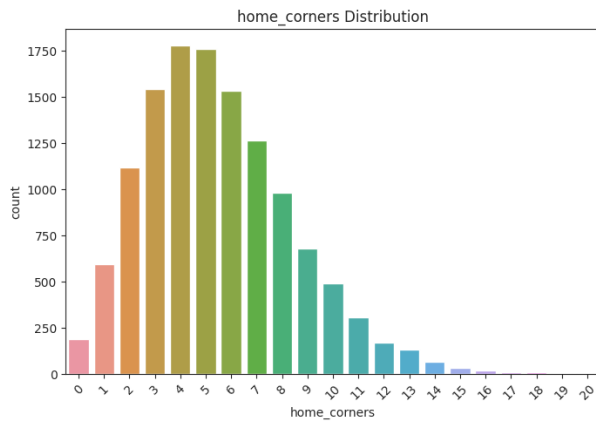
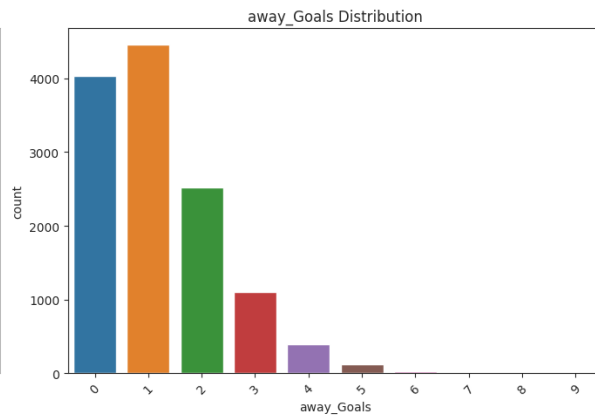
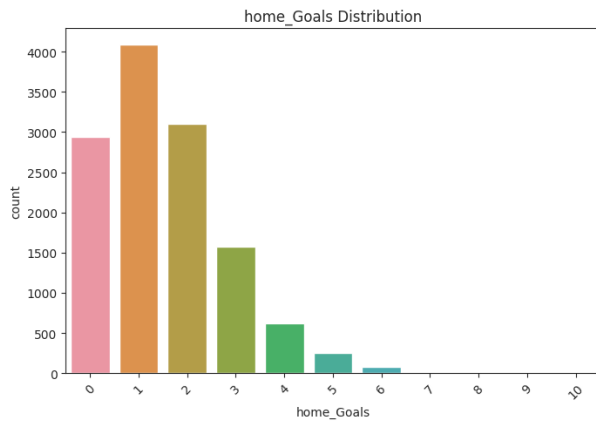
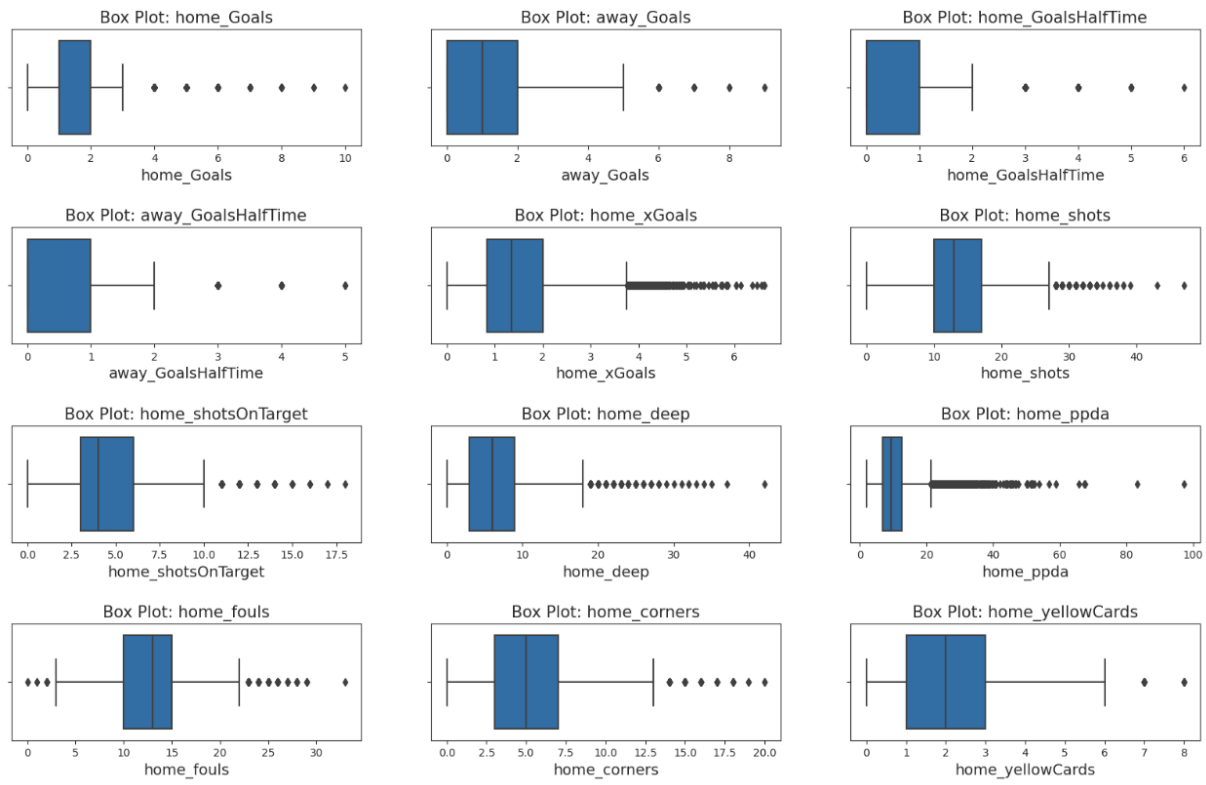
	playerID	teamID	playerName	teamName
0	560	89	Sergio Romero	Manchester United
1	557	89	Matteo Darmian	Manchester United
2	548	89	Daley Blind	Manchester United
3	628	89	Chris Smalling	Manchester United
4	1006	89	Luke Shaw	Manchester United
...
10101	7396	176	Loic Bessile	Bordeaux
10102	9566	175	Yanis Lhéry	Saint-Etienne
10103	9565	175	Mathys Saban	Saint-Etienne
10104	9568	181	Charles Costes	Dijon
10105	9567	181	Erwan Belhadji	Dijon

5. Methodology

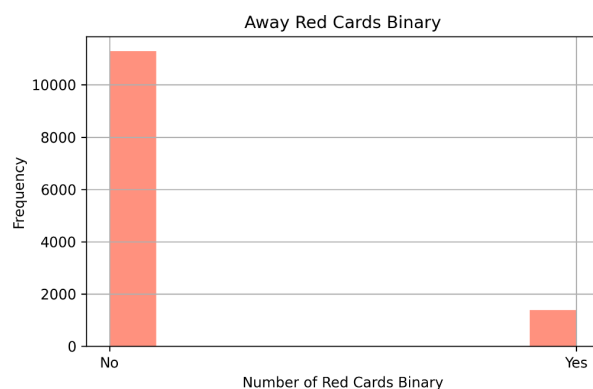
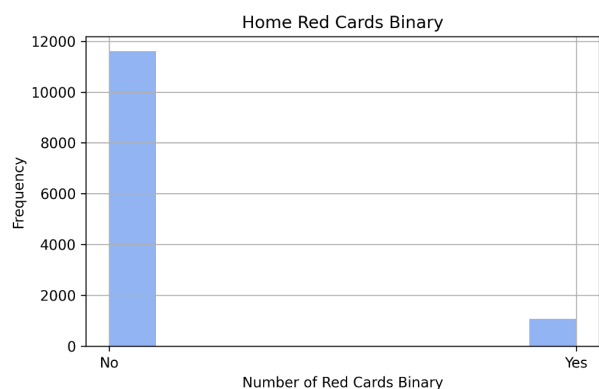
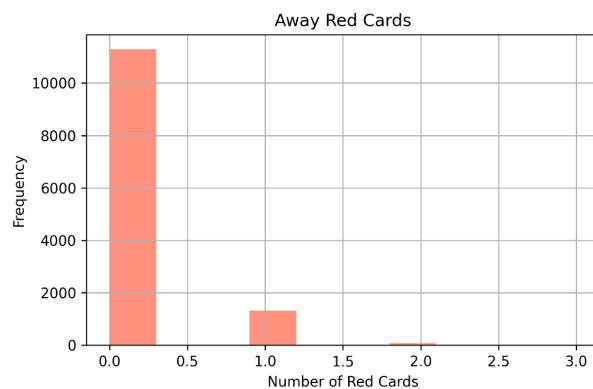
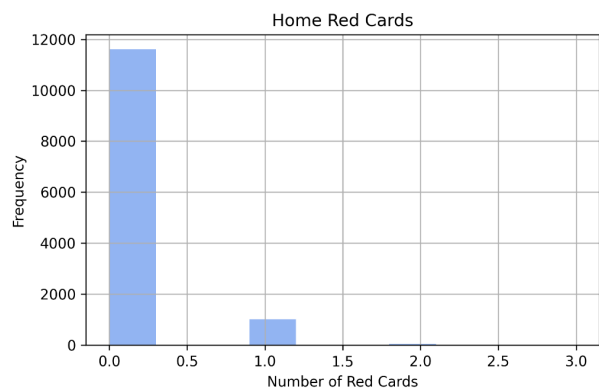
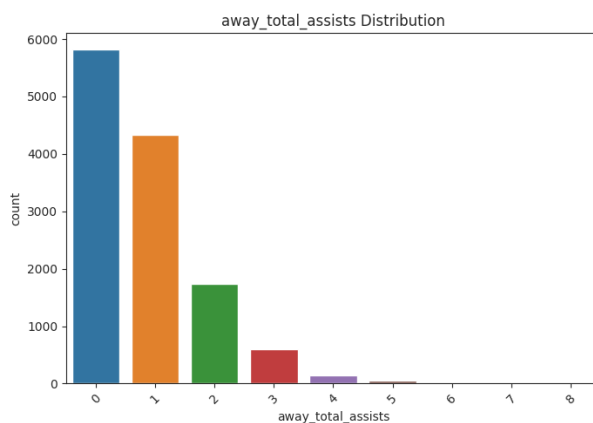
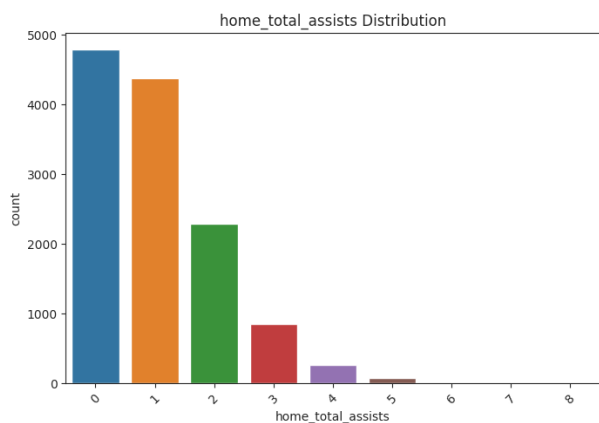
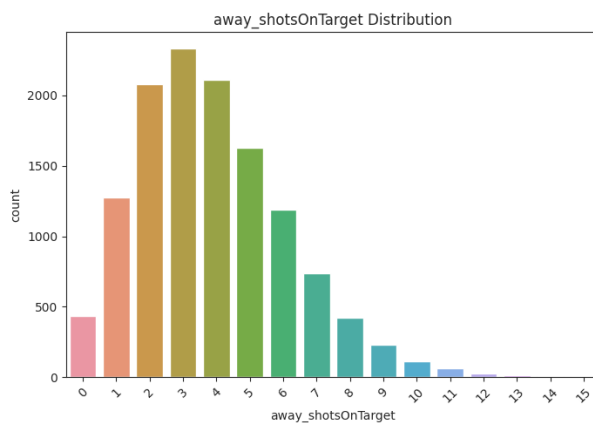
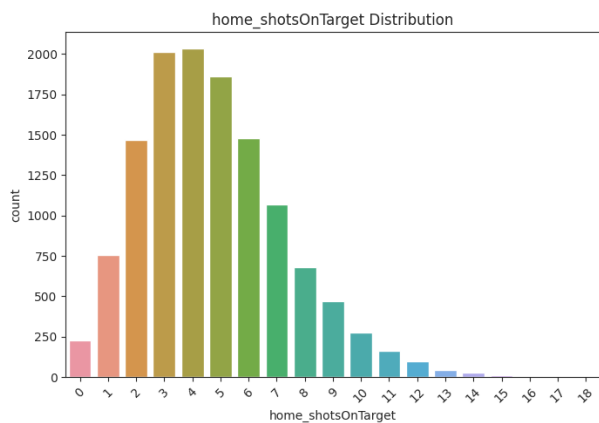
Includes data cleaning, EDA, feature engineering, feature selection, model training, and evaluation.

5.1 EDA

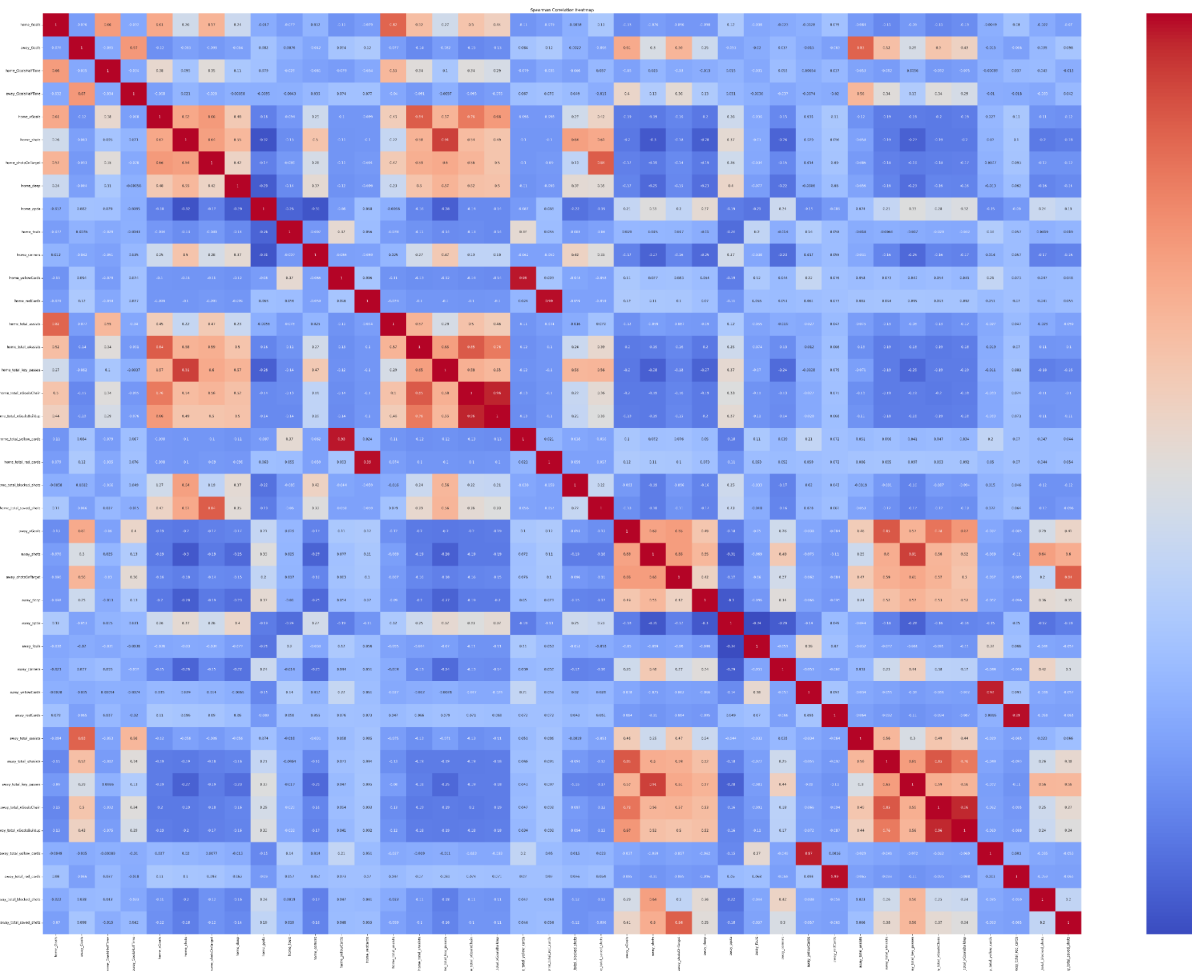
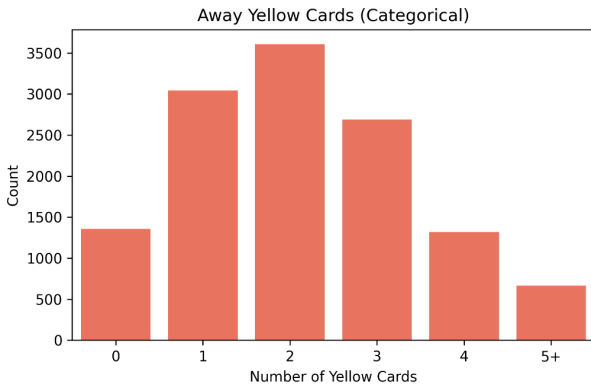
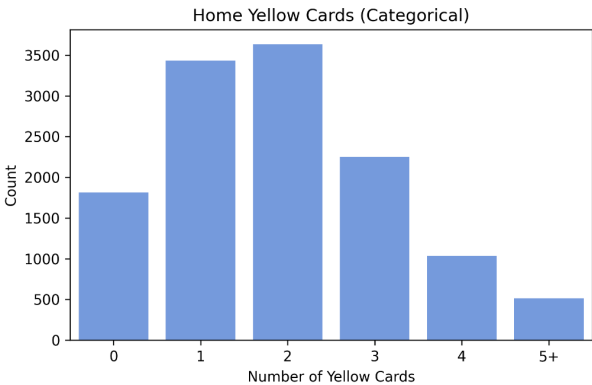
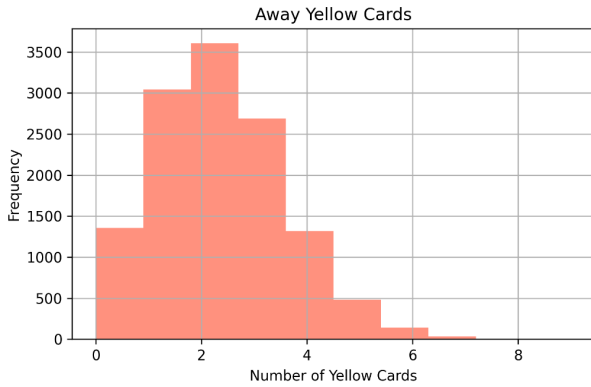
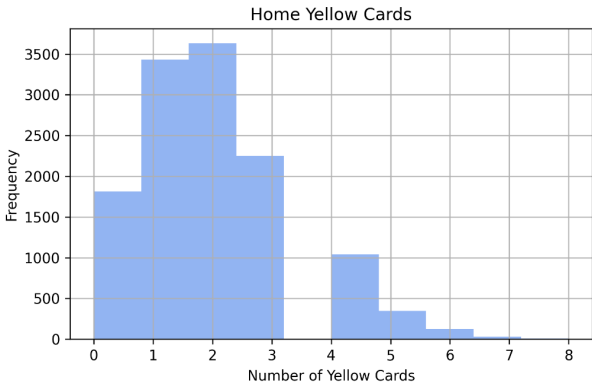
Football Match Result Prediction - Project Protocol



Football Match Result Prediction - Project Protocol

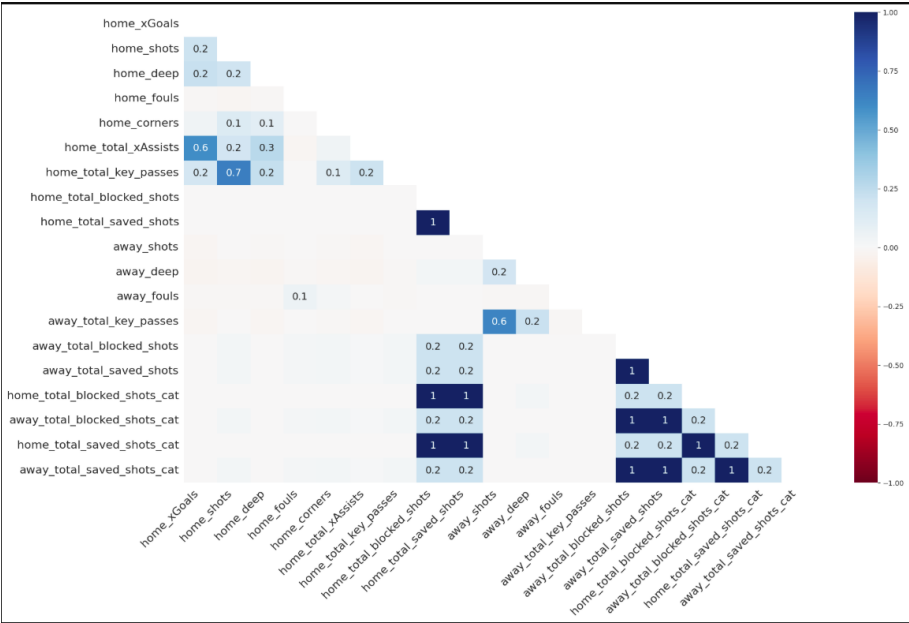


Football Match Result Prediction - Project Protocol

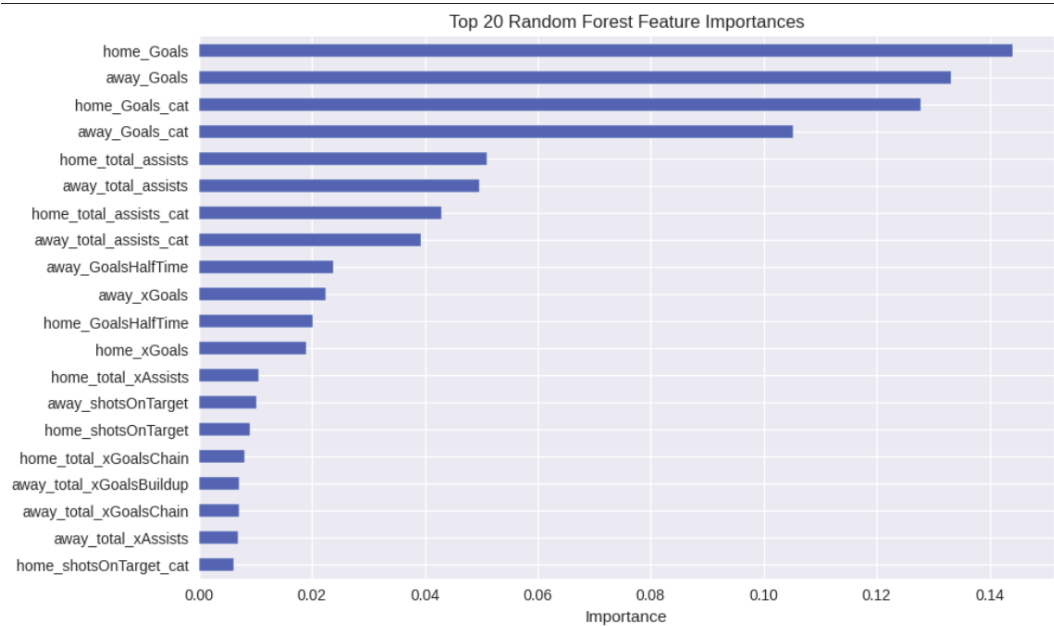


Football Match Result Prediction - Project Protocol

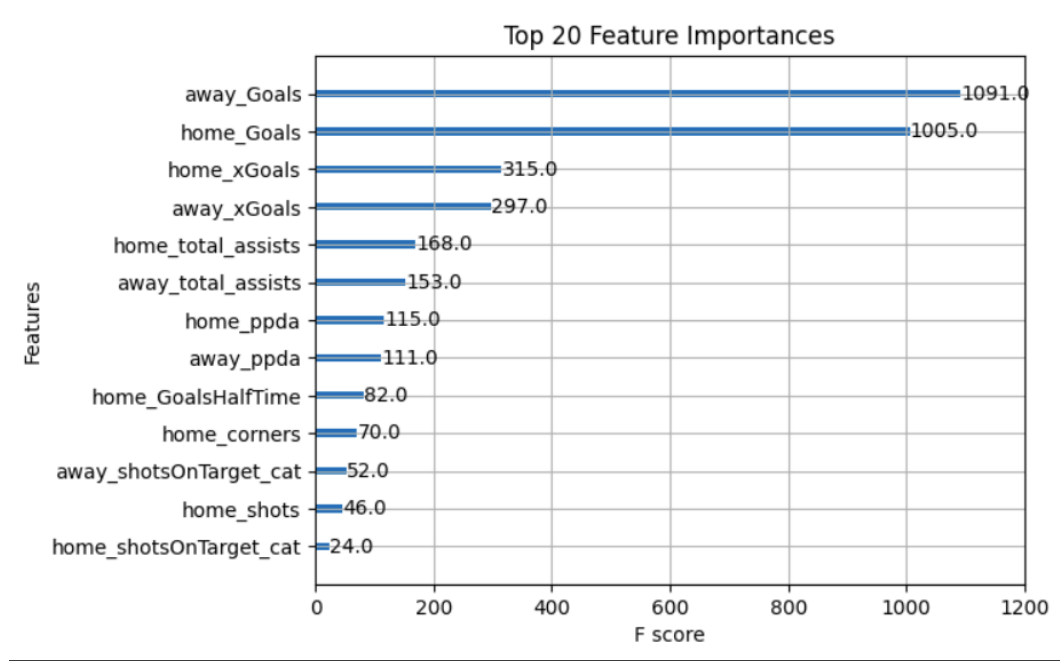
5.2 Missing Values



5.3 Feature Engineering & Selection



Football Match Result Prediction - Project Protocol



6. Model Training & Evaluation

Models tested: Logistic Regression, SVM, Decision Tree, Random Forest, GBM, AdaBoost, XGBoost.

Model	Accuracy	F1	LogLoss	AUC
Logistic Regression	1.000	1.000	0.0052	1.000
XGBoost	0.998	0.998	0.0052	1.000
GBM	0.999	0.999	0.0072	0.9999
Random Forest	0.998	0.998	0.0150	0.9999
SVM	0.994	0.994	0.0122	0.9999
Decision Tree	0.998	0.998	0.0710	0.9981
AdaBoost	0.662	0.681	0.6453	0.9631

7. Final Model Deployment

Final model: XGBoost Classifier with tuned hyperparameters, good performance on all sets.

Football Match Result Prediction - Project Protocol

8. Conclusion

A robust classifier was built with strong accuracy and potential real-world use.

Thank you,

Leonardo Romano